Appendix

Appendix A1 Study characteristics: Schirm, Stuart, & McKie, 2006 (randomized controlled trial with differential attrition)

Characteristic	Description
Study citations	Schirm, A., Stuart, E., & McKie, A. (2006). *The Quantum Opportunity Program demonstration: Final impacts.* Washington, DC: Mathematica Policy Research, Inc.[1] *Additional source:* Schirm, A., Rodriguez-Planas, N., Maxfield, M., & Tuttle, C. (2003). *The Quantum Opportunity Program demonstration: Short-term impacts.* Washington, DC: Mathematica Policy Research, Inc.
Participants	The *Quantum Opportunity Program* Demonstration Project used a randomized controlled trial research design. The demonstration operated in seven sites and served a single cohort of entering ninth graders over a five-year period. In six of the seven sites, the programs served ninth graders who entered high school in the fall of 1995. In one site (Washington, DC) the program served ninth graders who entered high school in the fall of 1996. *QOP* served students from high schools with dropout rates of 40% or more. To be eligible for *QOP*, students in these high schools had to meet the following three criteria: (1) they were entering ninth graders who were not repeating the ninth grade; (2) they had a grade point average below the 67th percentile of entering ninth graders at the participating high school; and (3) they did not have severe physical and learning disabilities that would prevent them from participating in the program. A sample of students meeting these criteria were drawn from lists of entering ninth graders; more than 97% of those identified agreed to participate in the study. The participating youth were then randomly assigned to either an intervention group that was enrolled in *QOP* or a control group that was not. Across the seven locations, 580 students were assigned to *QOP* group and 489 were assigned to the control group. Researchers compared the baseline characteristics of *QOP* and control group students on gender, age, race/ethnicity, and grade point average and found no statistically significant differences between the research groups. Participants were typically 13 or 14 years old, about two-thirds of participants were African-American, and a quarter were Hispanic. Participants were evenly split between males and females. Results summarized in this report are based on high school transcripts and three telephone surveys. One of these surveys was conducted at the end of the five-year demonstration, another two years after the demonstration had ended, and a third four years after the demonstration's end. There are two outcomes of interest for the WWC review of the effectiveness of *QOP*: total credits earned five years after program entry and high school diploma or GED certificate receipt within nine years of program entry. Total credits earned are based on transcript data and are available for 86% of *QOP* students and 77% of control group students. High school completion information is based on data from all three survey waves, as well as transcript data, and is available for 88% of *QOP* students and 83% of control group students.[2] For credits earned, the rate of differential attrition exceeded the 5% threshold used for WWC dropout prevention reviews. For high school completion, the rate of differential attrition was equal to this threshold. Because one measure used to rate *QOP's* effectiveness exceeded the differential attrition standard, the WWC rated this study as meeting evidence standards with reservations. To account for nonresponse, the study authors calculated impacts using weights that adjust for differences between respondents and nonrespondents in baseline characteristics. However, the WWC did not consider this statistical adjustment sufficient for overcoming the differential attrition.
Setting	This study took place at 11 high schools in seven evaluation sites: Fort Worth, Texas; Cleveland, Ohio; Washington, DC; Houston, Texas; Memphis, Tennessee; Philadelphia, Pennsylvania; and Yakima, Washington. Three sites implemented *QOP* in multiple high schools: Washington, DC (2 high schools), Houston (2 high schools), and Memphis (3 high schools). All other sites implemented *QOP* in a single high school.

(continued)

Appendix A1 Study characteristics: Schirm, Stuart, & McKie, 2006 (randomized controlled trial with differential attrition) *(continued)*

Characteristic	Description
Intervention	*QOP* Demonstration Project was an intensive, five-year, case management and mentoring program for high school youth that emphasized after-school supplemental education, developmental activities, and community service. Its primary goals were to increase the likelihood that enrollees would complete high school and enter a postsecondary education or training program. The program also aimed to reduce risky behaviors such as substance abuse, crime, and teenage parenting. *QOP* was operated by community-based organizations in seven sites. It offered a cohort of entering ninth graders services for up to five years and continued to provide services even if participants dropped out of school or moved out of the school district. The comprehensive program had four primary components: case management and mentoring, educational and developmental activities, supportive services, and financial incentives. These components are described in more detail below. 1. *Case management and mentoring:* Case managers typically had caseloads of 15 to 25 participants. Most case managers had office space within the school and were available to enrollees during the school day, at night, on weekends, and during school vacations. Case managers served as mentors to the enrollees and were typically assigned to the same enrollee for the full five years of the demonstration. According to an implementation study of *QOP,* most sites successfully implemented this component of the model (Maxfield et al., 2003). 2. *Education services, community service activities, and developmental activities: QOP's* participation target was 750 hours per year per enrollee. One-third of that time was to be spent on educational activities (such as tutoring or computer-assisted learning), one-third on community service (such as visiting residents of a local nursing home), and one-third on developmental activities (such as life skills and employment-readiness training). In most sites, based on program staff's assessment of participant need, resources were reallocated from community service activities toward case management and education activities. According to an implementation study of *QOP,* many sites did not fully implement *QOP's* education component, such as sustained course-based tutoring and computer-assisted instruction in basic reading and math skills (Maxfield et al. 2003). In addition, in many cases developmental activities designed to teach life skills were primarily recreational activities. Overall, enrollees spent an average of 174 hours per year on education, community service, and developmental activities—23% of the annual goal of 750 hours. Enrollees who spent little time participating in *QOP* activities frequently reported that they thought the program was too much like school or that they had a barrier to participation, such as a job or a child care or transportation problem (Maxfield et al., 2003). 3. *Supportive services: QOP* provided transportation assistance to facilitate attendance at program activities, as well as referrals to other resources in the community (such as mental health services and summer jobs programs). According to an implementation study of *QOP*, most sites successfully provided transportation services; however, most did not provide adequate child care support and did not consistently offer health screenings and referrals (Maxfield et al., 2003). *4.* *Financial Incentives: QOP* provided enrollees with a stipend of approximately $1.25 for every hour devoted to program activities other than mentoring or recreation. A matching amount was deposited in an accrual account, to be used by enrollees after they completed high school and enrolled in college, vocational training, or the military. Enrollees could also earn bonuses for achieving major milestones, such as a grade point average above a set benchmark. According to an implementation study of *QOP*, sites generally implemented this program component successfully (Maxfield et al., 2003).
Comparison	Control group members were not eligible to participate in *QOP* but could participate in other services available in the community. Based on responses to follow-up surveys, 16% of control group members participated in a program for disadvantaged youth other than *QOP* (Schirm et al., 2003). According to study authors, these other programs were generally substantially less intensive than *QOP.*

(continued)

Appendix A1 Study characteristics: Schirm, Stuart, & McKie, 2006 (randomized controlled trial with differential attrition) *(continued)*

Characteristic	Description
Primary outcomes and measurement	Two relevant outcomes from the *QOP* study are included in this summary and were used for rating purposes: the number of credits earned five years after program entry and high school diploma or GED certificate receipt within nine years of program entry. (See Appendices A2.1 and A2.2 for a more detailed description of outcome measures.) The study also examined the program's effects on academic outcomes, postsecondary outcomes, risky behaviors, resiliency factors, and attitudes. However, these outcomes do not fall within the three domains examined by the WWC's review of dropout prevention interventions (staying in school, progressing in school, and completing school). Therefore, these additional outcomes are not included in this report.
Staff training	*QOP* staff attended annual training conferences provided by Opportunities Industrialization Centers of America (OICA) during the demonstration period. The initial training lasted seven days with annual four-day sessions in subsequent years. The Ford Foundation funded technical assistance for all seven *QOP* demonstration sites to be delivered by OICA. Technical assistance activities included helping sites set up and maintain *QOP* management information systems, conducting site visits, helping resolve case management issues, and providing sites with developmental curriculum material and computer-assisted instruction (CAI) CD-ROMs. In addition, the U.S. Department of Labor provided technical assistance on selecting computer-assisted instruction (CAI) software, guidelines for setting up and operating accrual accounts, and quarterly calls with each site to discuss service delivery strategies.

1. The impact estimates summarized here come from two different reports from the same study. *QOP's* impact on credits earned used to rate its effectiveness in the progressing in school domain (and summarized in Appendix A3.1) was reported in Schirm, Rodriguez-Planas, Maxfield, & Tuttle (2003). High school completion impacts used by the WWC for rating *QOP's* effectiveness in the completing school domain (and summarized in Appendix A3.2) were reported in Schirm, Stuart, & McKie (2006).
2. This information was not available in the original report and was provided to the WWC by the study authors.

Appendix A2.1 Outcome measure for the progressing in school domain

Outcome measure	Description
Total credits earned by the end of year 5	This measure represents the total number of credits earned in the first five years after program entry. These data were collected through high school transcripts. These transcripts were requested from the high schools that students reported attending on follow-up surveys and are therefore only available for survey respondents. Credits are expressed in Carnegie units, in which one unit corresponds to a class that meets 45–60 minutes every day of the week for the entire academic year.

Appendix A2.2 Outcome measure for the completing school domain

Outcome measure	Description
Received high school diploma or GED by year 9	This binary measure represents the percentage of students who either received a high school diploma or a GED within nine years of program entry (five years after the expected graduation date for students making normal progress). This measure was based primarily on participants' response to the third survey. When this information was not available, it was supplemented with information from the first and second telephone surveys and transcript data.

Appendix A3.1 Summary of study findings included in the rating for the progressing in school domain[1]

Outcome measure	Study sample	Sample size[3] (students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *QOP* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*QOP* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Schirm, Stuart, & McKie, 2006 (randomized controlled trial with differential attrition)[8]								
Total credits earned by the end of year 5	Full sample	766	16.2 (9.0)	15.8 (8.5)	0.4	0.05	ns	+2
Domain average[9] for progressing in school						0.05	ns	+2

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement index in the progressing in school domain. These results were gathered from an earlier study report (Schirm et al. 2003).
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. Standard deviations for total credits earned were not included in the original study report and were provided to the WWC by the study authors.
3. Sample sizes for this measure were not included in the original study report and were provided to the WWC by the study authors. Transcripts were requested from the high schools that students reported attending on follow-up surveys. Therefore, this information is available only for those sample members who completed surveys. In addition, not all high schools provided the requested data, further reducing sample sizes for the credits earned measure.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Means from this study are weighted to adjust for differences between respondents and nonrespondents. In addition, each of the seven evaluation sites is weighted equally (and not weighted based on the number of participants in the site) in the calculation of these means.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. These results were gathered from an earlier study report (Schirm el al., 2003). The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In this case, no corrections for clustering or multiple comparisons were needed.
9. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the completing school domain[1]

Outcome measure	Study sample	Sample size[2] (students)	Authors' findings from the study: Mean outcome: *QOP* group	Authors' findings from the study: Mean outcome: Comparison group	WWC calculations: Mean difference[3] (*QOP* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Schirm, Stuart, & McKie, 2006 (randomized controlled trial with differential attrition)[7]								
Earned a high school diploma or GED by end of year 9 (%)	Full sample	915	78	75	3	0.10	ns	+4
Domain average[8] for completing school						0.10	ns	+4

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement index for the completing school domain. Appendix A4 reports the impact of *QOP* on earning a high school diploma, which is not used in *QOP*'s effectiveness rating.
2. Sample sizes for this measure were not included in the original study report and were provided to the WWC by the study's authors. This measure comes from the 791 respondents to the third follow-up survey. An additional 124 cases were added to this measure from those who did not respond to the third follow-up survey if their transcript or earlier survey responses indicated that they had earned a high school diploma or received a GED. This method makes it more likely to identify those who had completed high school or a GED by the end of year 9 than those who had not. To correct for this difference, the authors used nonresponse weights when estimating program impacts.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Means from this study are weighted to adjust for differences between respondents and nonrespondents. In addition, each of the seven evaluation sites is weighted equally (and not weighted based on the number of participants in the site) in the calculation of these means.
4. Effect sizes for dichotomous variables were computed using the Cox Index. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In this case, no corrections for clustering or multiple comparisons were needed.
8. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 Summary of additional findings for the completing school domain[1]

Outcome measure	Study sample	Sample size[2] (students)	Authors' findings from the study: Mean outcome: *QOP* group	Authors' findings from the study: Mean outcome: Comparison group	WWC calculations: Mean difference[3] (*QOP* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Schirm, Stuart, & McKie, 2006 (randomized controlled trial with differential attrition)[7]								
Earned a high school diploma by end of Year 9 (%)	Full sample	915	60	60	0	0	ns	0

ns = not statistically significant

1. This appendix presents effects of *QOP* on high school diploma receipt. The intervention's combined effect on high school diploma and GED receipt was used for determining the effectiveness rating and is presented in Appendix A3.2.
2. Sample sizes for this measure were not included in the original study report and were provided to the WWC by the study's authors. This measure comes from the 791 respondents to the third follow-up survey. An additional 124 cases were added to this measure from those who did not respond to the third follow-up survey if their transcript or earlier survey responses indicated that they had earned a high school diploma or received a GED. This method makes it more likely to identify those who had completed high school or a GED by the end of year 9 than those who had not. To correct for this difference, the authors used nonresponse weights when estimating program impacts.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Means from this study are weighted to adjust for differences between respondents and nonrespondents. In addition, each of the seven evaluation sites is weighted equally (and not weighted based on the number of participants in the site) in the calculation of these means.
4. Effect sizes for dichotomous variables were computed using the Cox Index. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In this case, no corrections for clustering or multiple comparisons were needed.

Appendix A5.1 *QOP* rating for the progressing in school domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of progressing in school, the WWC rated *QOP* as having no discernible effects. It did not meet the criteria for the other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, negative effects) because the one study that met WWC evidence standards did not show statistically significant or substantively important effects in this domain.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The one study of *QOP* passing evidence screens found no statistically significant or substantively important effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. *QOP* had only one study that met WWC evidence standards with reservations.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No statistically significant or substantively important negative effects were found in this domain.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. No statistically significant or substantively important positive effects were found in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No statistically significant or substantively important effects, either positive or negative, were found in this domain.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No statistically significant or substantively important effects, either positive or negative, were found in this domain.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No statistically significant or substantively important effects in this domain.

(continued)

Appendix A5.1 *QOP* rating for the progressing in school domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. No statistically significant or substantively important negative effects were found in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. No statistically significant or substantively important positive effects were found in this domain.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No statistically significant negative effects were found in this domain.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. No statistically significant or substantively important positive effects were found in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *QOP* rating for the completing school domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of completing school, the WWC rated *QOP* as having no discernible effects. It did not meet the criteria for the other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, negative effects) because the one study that met WWC evidence standards did not show statistically significant or substantively important effects in this domain.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The one study of *QOP* passing evidence screens found no statistically significant or substantively important effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. *QOP* had only one study that met WWC evidence standards with reservations.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No statistically significant or substantively important negative effects were found in this domain.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. No statistically significant or substantively important positive effects were found in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No statistically significant or substantively important effects, either positive or negative, were found in this domain.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No statistically significant or substantively important effects, either positive or negative, were found in this domain.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No statistically significant or substantively important effects in this domain.

(continued)

Appendix A5.2 *QOP* rating for the completing school domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. No statistically significant or substantively important negative effects were found in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. No statistically significant or substantively important positive effects were found in this domain.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No statistically significant negative effects were found in this domain.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. No statistically significant or substantively important positive effects were found in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Staying in school	na	na	na	na
Progressing in school	1	11	1,069	Small
Completing school	1	11	1,069	Small

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain, and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."